International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN (online): 2347-5552, Volume-11, Issue-2, March 2023
https://doi.org/10.55524/ijircst.2023.11.2.8
Article ID IRP1375, Pages 38-43
www.ijircst.org

A Review Article on the Prediction of Diseases at an Early Stage

Mohd Nadeem Khan[1], and Mrs. Ankita Srivastava[2]

[1]M.Tech. Scholar, Department of Computer Science and Engineering, Integral University, Lucknow, India
[2]Assistant Professor, Department of Computer Science and Engineering, Integral University Lucknow, India

ABSTRACT- Individuals today suffer from a wide range of diseases as a result of their lifestyle choices and the environment in which they live. The objective of forecasting disease at an earlier stage becomes an increasingly vital condition as the identification and prediction of such diseases at their earlier phases become highly significant. Most individuals are too lazy to go to the hospital or see a doctor for a small problem. Our approach focuses on accuracy to detect additional symptoms for illness prediction in healthcare. In this section, I've employed a variety of machine learning algorithms carefully and focused in this few, which achieved the highest accuracy with that specific condition in order to build a strong model that produces the most exact forecasts. This work introduces the topics of illness prediction, disease therapy, and local medical consultation with effective machine learning programming. There are several diseases in the world that are brought on by the conditions of people's living habits or their surroundings. Thus, this study offers a summary of machine learning-based illness prediction.

KEYWORDS- Disease Prediction, Machine Learning, Healthcare, Random Forest.

I. INTRODUCTION

Disease prediction is a method or approach for determining a person's health, whether he or she will have an illness or not, and to complete this ML model, it produces a prediction using the person's symptoms.All individuals nowadays deal with a variety of issues connected to chronic diseases as a result of unhygienic living practises, insufficient exercise, bad diets, and asymmetrical sleeping patterns. In order to prevent infections from spreading, early disease prediction is essential. The goal of ML is to simulate human cognitive processes. Due to the rapid advancement of analytics methods and the growing accessibility of medical data, this is bringing about a change in healthcare. The goal of this ML research in the healthcare industry is not to replace medical personnel but rather to function as an addition and help where humans fall short—specifically, swiftly processing huge volumes of data.ML is used in a wide range of industries in the modern world. Two examples are in the medical field and the education field.The availability of massive machine learning's application in the real world has grown in response to greater datasets and computational capacity. Disease Prediction is a web application that customers may use on their phones or computers to monitor and visualize the condition they are suffering from. The healthcare industry has used ML in a variety of ways. The abundance of images, patient information, and other information makes it easier to spot trends and, as a consequence, create

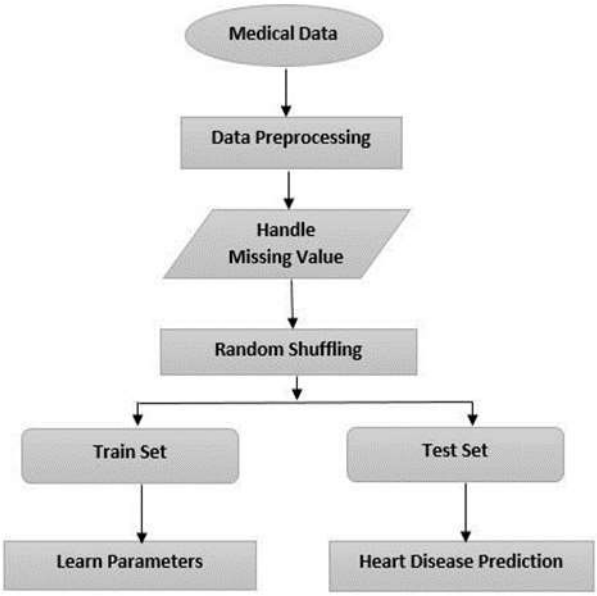

Figure 1: Analysing Model of Disease Prediction

algorithm-based projections. Machine learning algorithms, as we all know, employ a variety of a experimental and numerical methods but also build methodologies that understand information produced by previous events and apply it in higher mental abilities. Heart attacks have also been reported as a result of the worldwide pandemic coronavirus. COVID-19 can induce cardiac muscle edema. As a result of the high mental stress caused by COVID-19, heart disease is becoming increasingly widespread in metropolitan areas. A sickness Based on the test results, a prediction model informs the individual if they have a certain condition or not. A disease prediction algorithm may detect when an individual should have any illness without any physical contact in scenarios such as COVID-19 and statewide shut downs, in which individuals cannot contact physicians or commute to hospitals. ML models come in a variety of forms, including supervised, semi-supervised, but also unsupervised learning techniques, as well as deep learning or reinforcement learning methodologies. These understandings are applied to quickly categorise huge amounts of data. We use ML methods like DT, RF, NB classifiers to accomplish quick categorization of vast volumes of data and accurate illness

prediction. With the continuous expansion of medical data, we were able to extract the necessary information to anticipate the condition. Data science as well as big data can be utilised for the prediction of several types of diseases using past health data obtained from the patient. These illness prediction models are critical to detecting the presence of illness.

II. COMPARATIVE STUDY

Sr. No.	Author	Title of the Paper	Purpose of the Paper	Technique(s) Used & Accuracy	Research Gap
1	**Rohit Kumar et al[1]**	Special Predictionof disease Using ML	The paper's objective is to provide a solution to the problem of early detection and accurate diagnosis of special diseases, which could help improve patient outcomes and reduce healthcare costs.	Naïve Bayes classifier Algorithm. 78.6%	Future work could include expanding the scope to include other diseases or conditions, such as neurological disorders or mental health conditions.
2	**Adarsh Kumar Pandey et al[2]**	Disease Prediction Using ML Algorithms	The paper's objective is demonstrate the potential of a machine learning in improving healthcare outcomes by enabling early detection and prevention of diseases.	DT Algorithm, RF Algorithm, NB Algorithm. 95%	Future work could involve comparing additional algorithms to identify the most suitable algorithm for disease prediction.
3	**Dhriti Gada et al[3]**	Disease Prediction System using ML	The paper's objective is provide practical solution for healthcare professionals to predict and diagnose diseases accurately and promptly.	KNN, Logistic Regression,DT Classifier, and RF Classifier. 96.45%	Future work should be Developing an interface that would allow doctors and healthcare professionals to input patient data and receive disease predictions in real-time.
4	**Abhimanyu Barun et al[4]**	Various Disease Predictions By the use of ML	The paper's objective is provide insights into the suitability of different ML models for the various types of diseases and to highlight the potential benefits of early disease prediction, such as improved prevention and treatment outcomes.	Logistic Regression,Ada Boost Classifier,Naive Bayes classifier,XGB Classifier,Gradient Boosting Classifier.97.30 %	The author suggests that future research look at integrating machine learning models into medical decision-support systems that help doctors make informed decisions regarding illness prevention and treatment.
5	**R.M. Gomathi et al[5]**	A Perfect Multi-Perspective View for Disease Prediction Using a Machine Learning Method	The goal of this work will be to help create more precise and complete illness prediction models through the use of machine learning techniques, with the potential to enhance patient outcomes and guide tailored healthcare treatments.	DT Algorithm, RF Algorithm, NB Algorithm. 92-95%	The author suggests that future research is related to the need for more accurate and comprehensive disease prediction models that take into account multiple perspectives or sources of information.
6	**Animesh Basak et al[6]**	Detection of Heart Problems Using a Machine Learning Method	The paper's objective is development a cost-effective as well as non-invasive method for detecting and predicting heart disease, which can have significant implications for improving public health.	SVM, KNN, NB, and RF. 93%	Research gap in this paper can be considered as the need to develop a predictive model on heart problem using larger and more diverse dataset, which includes a wide range of medical and demographic features.
7	**Anshika Trisal et al[7]**	Cardiac Disease Prediction with ML Algorithms	The goal of this research is to illustrate the capability of machine learning in predicting heart illness and to emphasise the relevance of applying such technologies in healthcare to improve patient outcomes.	KNN. DT, SVM. 98%.	Future research could explore the generalizability of the findings in this paper by using a more diverse dataset and analyzing the influence of different patient characteristics on the accuracy of the machine learning algorithms in predicting cardiac disease.
8	**Ayushi Sharma**	Disease Prediction	The paper's objective is demonstrate the potential for	KNN, NB,	There may be a need for a more in-depth analysis that considers

	et al[8]	using machine learning algorithms	machine learning algorithms to predict diseases.	SVM, DT. 94%	multiple factors, such as the type of disease, the size and quality of the dataset, and the specific machine learning algorithm used.
9	**Monali Gulhane et al[9]**	A disease prediction model based on machine learning	The paper's objective is develop accurate as well as efficient disease prediction models that can aid in early diagnosis and better patient outcomes.	Convolutional neural networks (CNNs), K-Nearest Neighbor (KNN) 98.7%	Future work could explore additional risk factors such as genetics, environmental exposure, or socio-economic factors.
10	**Md. Ehtisham Farooqui, et al[10]**	A Comprehensive Analysis of Machine Learning-Based Disease Prediction Models	This study paper's goal is to offer a thorough analysis of several machine learning-based illness prediction models.	NB, RF, KNN, DT, SVM. 94.8%	The paper's research need might be a more in-depth investigation of the efficacy of machine learning models in predicting certain illnesses.
11	**Nishant, Ritik, et al[11]**	Disease Prediction by Machine Learning Over Big Data from Healthcare Communities	The goal of the study article is to contribute to the creation of illness prediction models that are more precise and effective, which can enhance patient care and healthcare outcomes.	Machine Learning, SVM Algorithm, K-Mean. 94.8%	The paper mainly focuses on the technical aspects of developing a disease prediction model using machine learning algorithms, but it does not address the ethical and legal issues that may arise when using sensitive patient data for healthcare research.
12	**Sneha Grampurohit, et al[12]**	Machine Learning Algorithms for the Disease Prediction	The goal of the research study is to create a disease prediction model that, using a patient's medical history and demographic data, can reliably forecast the chance that a patient would contract a certain illness.	NB, RF, KNN, DT, SVM classifier. 95%	This study uses a single dataset for training and testing the disease prediction models. Future research could validate the proposed model on external datasets to determine its generalizability and robustness.
13	**Kedar Pingale, et al[13]**	Disease Prediction with the use of Machine Learning	The objective of the study is to create a machine learning-based illness prediction model that can precisely forecast the likelihood of diseases based on numerous patient characteristics.	CNN-MDRP, K-nearest Neighbors, Decision Tree, Naïve Bayesian. 94.8%	The study uses a relatively small dataset, which may not be representative of the entire population. Future studies might assess the generalizability of the suggested model using bigger and more varied datasets.
14	**Akash C, et al[14]**	Disease Prediction Using Big Data from Healthcare Communities	Research paper's objective The study uses the vast volumes of data produced from several sources, including electronic health records, medical sensors, and patient-reported outcomes, to solve the difficulty of illness prediction, which may be difficult and time-consuming.	(CNN-MDRP), NB Algorithm, DT algorithm. 95%	Future research could focus on developing methods to improve the interpretability of machine learning models for disease prediction. This could include techniques such as feature importance analysis, decision tree visualization.
15	**Saiesh Jadhav et al[15]**	Disease Prediction with the use of Machine Learning from Healthcare Communities	The goal of the research article is to create a disease prediction model based on machine learning utilising information gathered from healthcare communities.	KNN algorithm CNN-UDRP Support Vector Machine (SVM). 94.8%	The study suggests that future research should focus on developing more sophisticated data standardization and privacy techniques to ensure the ethical use of health data.
16	**Dhiraj Dahiwade et al[16]**	Creating a Model for Illness Prediction Using Machine Learning	The goal of the study article is to use the vast volumes of data collected from several sources, including electronic health records, medical sensors, and patient-reported outcomes, to solve the difficulty of illness prediction, which may be difficult and time-consuming.	K-Nearest Neighbor (KNN), Convolutional neural network (CNN). 84.5%	The study also proposes that to increase the precision and interpretability of illness prediction models, future research should concentrate on creating more advanced feature selection algorithms.

17	**Vinitha S et al[17]**	DISEASE PREDICTION USING MACHINE LEARNING OVER BIG DATA	In order to forecast disease and identify risk factors, the research study provides a framework that combines several machine learning techniques, such as Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Artificial Neural Networks (ANN).	DT, RF, SVM, and ANN. 94.8%	According to the report, future research should concentrate on creating machine learning models that are easier to understand and that can offer valuable information for clinical decision-making.
18	**Tanner Christensen et al[18]**	Machine Learning Techniques in Predicting Illness Using Claims Data	The goal of the research article is to investigate how claims data may be used to use machine learning algorithms for illness prediction.	Logistic Regression (LR), Random Forest (RF), XGBoost (XGB), LSTM. 96.4%	The study suggests that future research should explore the integration of claims data with other sources of healthcare data, such as electronic health records, medical sensors, and patient-reported outcomes, to develop more comprehensive and accurate disease prediction models.
19	**Shraddha Subhash et al[19]**	Big Data and Machine Learning for Disease Prediction	The goal of the study article is to increase illness prediction's accuracy and effectiveness, which might result in earlier diagnosis, more successful treatment, and improved patient outcomes.	CNN-UDRP algorithm, Naive Bayes, Decision tree algorithm , SVM. 94.8%	The limited availability of quality healthcare data. The study acknowledges that while big data holds immense potential for disease prediction, the quality and accuracy of the data are crucial for the success of machine learning models.
20	**MIN CHEN et al[20]**	Machine learning-based disease prediction using large-scale healthcare community data.	The goal of the study paper is to look at how large data from the healthcare industry might be used to forecast diseases using machine learning techniques.	DT, RF, SVM, and ANN Classifier. 94.8%	The study emphasizes that the quality and accuracy of health data are crucial for the success of machine learning-based disease prediction models. However, due to the heterogeneity and complexity of health data sources, it is often difficult to ensure the consistency and reliability of the data.

III. FINDING AND DISCUSSION

In this section, we provide the process of determining the presence of the disease in a patient according to the symptoms they have. So, under this architecture, we use many ML classification techniques like RF, SVM, Decision Tree, Naive Bayes, etc. to classify the presence of the disease on a patient's body.

- Disease prediction using machine learning seems to have the chance to enhance patient experiences and lower medical expenses by enabling early identification and intervention.
- Gathering relevant patient data from electronic health records (EHRs) and developing prediction models are among the most common methods to forecast sickness using ML. These models might be used to identify patients who are at high risk of developing an illness, allowing healthcare providers to intervene before the ailment progresses.
- A range of maladies, including cardiovascular disease, mental health difficulties, chronic renal disease, Alzheimer's disease, and Parkinson's disease, have been treated using machine learning algorithms.
- ML models function differently depending on the disease being examined and the data provided. In some situations, machine learning models have predicted illness outcomes with great accuracy, while in others, the performance has been relatively modest.
- The necessity for vast volumes of high-quality data is one of the problems of illness prediction using machine learning. This is especially challenging for uncommon illnesses or those with extensive latency periods.
- Machine learning model interpretability and transparency are essential factors since they can help develop trust in the models and promote their adoption by healthcare professionals.
- Lastly, when utilizing machine learning for illness prediction, there are key ethical factors to consider, such as privacy, bias, and informed consent.
- In above discussion, illness prediction using machine learning has enormous promise for improving healthcare outcomes, but it also introduces a number of problems and ethical concerns that must be addressed. Further research is needed to improve machine learning models' performance and interpretability, as well as to guarantee that they are used responsibly and ethically**.**

IV. CONCLUSION

Based on the findings of the 20 studies on illness prediction using machine learning, it is possible to infer that machine learning has considerable potential for improving disease prediction accuracy and outcomes in a variety of medical disorders. The experiments demonstrated that machine learning models may extract significant elements from a variety of data sources, including EEG signals, electronic

health records, gene expression data,and others, to forecast illness progression or risk.Despite the positive results, the research revealed various obstacles, including data quality, interpretability, and model generalizability. Furthermore, the majority of research used small datasets or was confined to specific communities, limiting the generalizability of findings to larger and more varied populations.There are various potential future study areas based on the findings of these studies. Secondly, future research might concentrate on building more interpretable machine learning models that can give insights into the variables driving predictions, therefore assisting clinical decision-making. Second, to increase model generalizability, the datasets employed in this research should be expanded to encompass bigger and more varied populations. Lastly, combining several data sources can increase forecast accuracy even further. Fourth, building models that can predict illness development in real time can enable early intervention and enhance patient outcomes. Lastly, In future work we can say that the above comparative study has revealed that the most experiments was performed through combining ml techniques moreover throughout order to enhance a models performances we do have to enhance the quality of a datasets additionally new features could be introduced to improve existing results.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Kumar, R., Thakur, P., & Chauhan, S. (2022). Special Disease Prediction System Using Machine Learning. 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, COM-IT-CON 2022, 42–45. https://doi.org/10.1109/COM-IT-CON54601.2022.9850843.

[2] Pandey, A. K., Tripathi, A., & Ranjan, R. (2022). Disease Prediction Using Machine Learning (Health Buddy). 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022, 1003–1006. https://doi.org/10.1109/ICACITE53722.2022.9823602.

[3] Gada, D. (2022). Disease Prediction System using Machine Learning. 2022 6th International Conference on Computing, Communication, Control and Automation, ICCUBEA 2022. https://doi.org/10.1109/ICCUBEA54992.2022.10011002.

[4] Barun, A., Nanda, M., Jarwal, M. K., & Sipal, D. (2022). Various Disease Predictions Using Machine Learning. 2022 IEEE Region 10 Symposium, TENSYMP 2022. https://doi.org/10.1109/TENSYMP54529.2022.9864521.

[5] Gomathi, R. M., Deepa Jothi, K., Ajitha, P., Sivasangari, A., Anandhi, T., & Rani, V. N. (2022). Flawless Multi Perspective Vision for Prediction of Disease using Machine Learning Approach. 2022 6th International Conference on Trends in Electronics and Informatics, ICOEI 2022 - Proceedings, 1190–1194. https://doi.org/10.1109/ICOEI53556.2022.9776787.

[6] Basak, A., Rahman, M. S., & Rahman, M. (2022). Prediction of Heart Disease Using an Approach Based on Machine Learning. 2022 13th International Conference on Computing Communication and Networking Technologies, ICCCNT 2022. https://doi.org/10.1109/ICCCNT54827.2022.9984555.

[7] Trisal, A., Sagar, V., & Jameel, R. (2022). Cardiac Disease Prediction using Machine Learning Algorithms. Proceedings of International Conference on Computational Intelligence and Sustainable Engineering Solution, CISES 2022, 583–589. https://doi.org/10.1109/CISES54857.2022.9844370.

[8] Sharma, A., Pathak, J., & Rajakumar, P. (2022). Disease Prediction using machine learning algorithms. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022, 995–999. https://doi.org/10.1109/ICACITE53722.2022.9823744.

[9] Gulhane, M., & Sajana, T. (2021). A Machine Learning based Model for Disease Prediction. 2021 International Conference on Computing, Communication and Green Engineering, CCGE 2021. https://doi.org/10.1109/CCGE50943.2021.9776374.

[10] Farooqui, Md. E., & Ahmad, Dr. J. (2020). A DETAILED REVIEW ON DISEASE PREDICTION MODELS THAT USES MACHINE LEARNING. International Journal of Innovative Research in Computer Science & Technology, 8(4). https://doi.org/10.21276/ijircst.2020.8.4.14.

[11] Usman, A., & Rajpurohit, B. S. (2021). Modeling and classification of stator inter-turn fault and demagnetization effects in bldc motor using rotor back-emf and radial magnetic flux analysis. International Journal of Power and Energy Systems, 41(2). https://doi.org/10.1109/ACCESS.2017.

[12] Jain College of Engineering, Institute of Electrical and Electronics Engineers. Bangalore Section., & Institute of Electrical and Electronics Engineers. (n.d.). 2020 International Conference for Emerging Technology (INCET) : Belgaum, India. Jun 5-7, 2020.

[13] Pingale, K., Surwase, S., Kulkarni, V., Sarage, S., & Karve, A. (2019). Disease Prediction using Machine Learning. International Research Journal of Engineering and Technology. www.irjet.net.

[14] Jamgade, A. C. (2019). Disease Prediction Using Big Data from Healthcare Communities. International Journal of Advanced Research in Computer and Communication Engineering, 8 https://doi.org/10.17148/IJARCCE.2019.8328.

[15] Jadhav, S., Kasar, R., Lade, N., Patil, M., & Kolte, S. (2019). Disease Prediction by Machine Learning from Healthcare Communities. International Journal of Scientific Research in Science and Technology, 29–35. https://doi.org/10.32628/ijsrst19633.

[16] Surya Engineering College, & Institute of Electrical and Electronics Engineers. (n.d.). Proceedings of the 3rd International Conference on Computing Methodologies and Communication (ICCMC 2019) : 27-29, March 2019.

[17] Vasumathi, M. T., & Kamarasan, M. (2019). Fruit disease prediction using machine learning over big data. International Journal of Recent Technology and Engineering, 7(6), 556–559. https://doi.org/10.5121/cseij.2018.8101.

[18] Christensen, T., Frandsen, A., Glazier, S., Humpherys, J., & Kartchner, D. (2018). Machine learning methods for disease prediction with claims data. Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018, 467–471. https://doi.org/10.1109/ICHI.2018.00108.

[19] Subhash Shirsath, S., & Patil, S. (2018). Engineering and Technology (A High Impact Factor. International Journal of Innovative Research in Science, 7. https://doi.org/10.15680/IJIRSET.2018.0706059.

[20] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease Prediction by Machine Learning over Big Data from Healthcare Communities. IEEE Access, 5, 8869–8879. https://doi.org/10.1109/ACCESS.2017.2694446.

ABOUT THE AUTHORS

Mohd Nadeem Khan is M.Tech Scholar in Department of Computer Science and Engineering from Integral University, Uttar Pradesh, Lucknow, India. His researth area is machine learning. His email id is nadeemk0076@gmail.com

Ankita Srivastava is an Assistant Professor in Department of Computer Science & Engineering from Integral University, Uttar Pradesh, Lucknow.She has published many research paper in different different International journal and conferences.